# Applying Spatial Autocorrelation Analysis to Evaluate Error in New England Forest-Cover-Type Maps Derived from Landsat Thematic Mapper Data

**Scott A. Pugh and Russell G. Congalton**

## Abstract

*A spatial autocorrelation of error analysis was performed to compare the patterns of error in a classified Landsat Thematic Mapper (TM) forest-cover-type image for two forested areas—one public and one private. TM data were classified to generate a detailed forest type map, and intensive ground reference data covering approximately 3600 ha were collected for both study areas. Two difference images were produced by comparing the reference inventory with the classified data, pixel by pixel. The subsequent spatial autocorrelation analysis indicated that concentrated blocks of error were more pronounced in the public lands study area than in the private lands study area, where error was more evenly distributed. The results indicated that systematic sampling is not always suitable for assessing error in TM data.*

## Introduction

Assuming mutually exclusive units in a two-dimensional plane, spatial autocorrelation exists if the presence, absence, or degree of a certain characteristic affects the presence, absence, or degree of the same characteristic in neighboring units (Cliff and Ord, 1981). Because most natural phenomena are interdependent and distributed geographically in patches or gradients (Legendre, 1993), spatial autocorrelation is an important factor in remote sensing studies of the environment and natural resources.

Remotely sensed data exhibit a dependency between neighboring pixels. Previous research has shown that spectral reflectance values and classification error among neighboring pixels tend to be interdependent in remotely sensed data (Tubbs and Coberly, 1978; Craig and Labovitz, 1980; Campbell, 1981; Labovitz *et al.*, 1982; Congalton, 1988a). Consequently, spatial autocorrelation affects many aspects of remotely sensed data, including results of preprocessing and classifying algorithms as well as sampling designs and hypothesis testing for accuracy assessment. For example, samples acquired in close proximity to each other from populations distributed in patches or gradients tend to have similar values and thus exhibit positive spatial autocorrelation. Negative spatial autocorrelation—less common in environmental phenomena—occurs when values tend to be dissimilar.

Autocorrelation in remotely sensed data can be influenced by a number of factors, including atmospheric haze, physiography, and cover-type distribution. Autocorrelation of spectral reflectance values was detected in Landsat Multispectral Scanner (MSS) data (Craig and Labovitz, 1980; Campbell, 1981; Congalton, 1988a). Craig and Labovitz (1980) determined cloud cover and geographic location (e.g., ocean versus land, forest versus tundra) to contribute to this autocorrelation. Labovitz *et al.* (1982) also detected autocorrelation of spectral reflectance values in Landsat Thematic Mapper Simulated (TMS) data and discovered more evidence linking autocorrelation with atmospheric haze. They further determined that physiography was more important than land-cover type in explaining the geographic location effect on autocorrelation of TMS spectral reflectance values. Campbell (1981) observed autocorrelation of MSS spectral reflectance values in homogeneously forested areas (i.e., clusters of pixels with similar characteristics).

In addition to physiography, atmospheric haze, and cover type distribution, preprocessing techniques and imaging spectrometer properties may also cause autocorrelation. Congalton (1988a) determined, for example, that resampling may result in spatial autocorrelation of error in remotely sensed data. In resampling—a preprocessing technique in which the remotely sensed image is geometrically corrected—the new locations of the transformed (resampled) pixels usually do not coincide with the locations of the source pixels. The digital numbers of the transformed pixels must therefore be estimated from the digital numbers of neighboring pixels, which may introduce spatial autocorrelation.

Sensor spatial resolution may also influence spatial autocorrelation. Labovitz *et al.* (1982) investigated the effects of resolution on the autocorrelation of spectral reflectance values using sequential autocorrelation statistics. They compared MSS scan line reflectance values with higher spatial resolution TMS scan line reflectance values of comparable wavelength from equivalent regions. The spectral reflectance values of the higher resolution TMS sensor (7.5- to 18.39-m resolution) were more autocorrelated than the spectral reflectance values of the MSS sensor (80-m resolution). Labovitz *et al.* (1982) suspected that this greater magnitude of autocorrelation was a result of the number of spectrally homogeneous areas larger than the TMS spatial resolution as compared to the number of spectrally homogeneous areas larger than the MSS spatial resolution. Furthermore, the lengths of the slopes in the study area were occu-

Department of Natural Resources, University of New Hampshire, Durham, NH 03824 (russ.congalton@unh.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 67, No. 5, May 2001, pp. 613–620.

0099-1112/01/6705–613$3.00/0

pied by more TMS pixels than MSS pixels. Differences between the sensors other than spatial resolution, however, were thought to be minimal.

Campbell (1981) and Congalton (1988a) observed spatial autocorrelation of error in classified MSS data. By using spatial autocorrelation statistics, Campbell (1981) demonstrated that classification errors were clustered. In other words, the error was not randomly distributed. In addition, he found that autocorrelation could change with temporal scene variability. He determined that the degree of positive continuity of MSS spectral values among adjacent pixels varied from date to date in the same year. Because the degree of spectral autocorrelation varied throughout the year, the amount of classification error also varied. While Campbell's (1981) small data set may have been insufficient to make generalizations, his research suggested that choosing a sampling scheme to assess error without considering spatial autocorrelation present in the data set could bias estimates of error.

Congalton (1988a; 1988b) performed extensive research investigating the effects of autocorrelation on accuracy assessments in classified MSS data (50- to 63.6-m resolution). The effects of autocorrelation on different sampling schemes for three equal-size study areas of varying spatial complexity were investigated. The results showed that the best sampling schemes corresponded with particular types of spatial autocorrelation. For example, systematic and stratified systematic unaligned sampling did not perform well when the periodicity of error (spatial autocorrelation) was extreme. In contrast, simple random sampling performed well over all degrees of periodicity, but worked best when the periodicity was extreme. Stratified random sampling performed best when the periodicity of error was moderate, and cluster sampling provided suitable estimates for all degrees of periodicity.

Spatial autocorrelation of error may also be affected by classification techniques. For example, specific cover types may be consistently classified incorrectly with particular methods (e.g., inadequate training area selection). If the cover types occur in specific patterns throughout the image, then error will follow this pattern. The classification process, especially the selection of training areas, should therefore be carefully evaluated when assessing classification error. Nevertheless, image classification itself is a culmination of many processes, each of which can affect spatial autocorrelation of error.

There is still much to be learned about spatial autocorrelation of error. Congalton (1988a) stressed that continued research focused on understanding spatial autocorrelation of error in higher resolution imagery is necessary to fully comprehend and utilize information in remotely sensed data. To that end, this project compared the spatial autocorrelation of error in classified Landsat TM imagery (28.5-m resolution) for two forested environments in New Hampshire, one public and one private. The main objectives were as follows:

- to quantify the magnitude and pattern of error in the public and private lands study areas, and
- to compare the study areas to investigate causes of error and determine if differences between the study areas cause spatial autocorrelation of error to differ.

## Methods

### Spatial Autocorrelation Statistics

In area-based data, such as remotely sensed data, an area is partitioned into $n$ non-overlapping regions. Each region is associated with a variable or characteristic, $x$. Spatial autocorrelation statistics describe the spatial arrangement of this variable or characteristic of interest, and may yield information which explains the underlying phenomena causing the observed pattern. The characteristic of interest can be continuous or discrete. The simplest type of discrete data is binary data, which indicates only whether a certain characteristic, $x$, is present or not. Error in remotely sensed data is binary; the data are either correctly classified or they are not.

There are several commonly used spatial autocorrelation statistics, including the join count statistics, Moran's $I$ statistic, and the Geary's $c$ statistic. While all of these statistics can be used with discrete data, the join count statistics are the simplest and most widely employed. Moran (1948) described join count statistics in accordance with binary data: i.e.,

Define the characteristic of interest, $x$, as follows:
$x_i = 1$ if the $i$th region possesses the characteristic
$x_i = 0$ if the $i$th region does not possess the characteristic
Define the join matrix, $\{\delta_{ij}\}$, as follows:
$\delta_{ij} = 1$ if the regions $i$ and $j$ have a common boundary of positive length
$\delta_{ij} = 0$ if the regions $i$ and $j$ do not have a common boundary of positive length

The total number of observed 1-1 joins (i.e., the number of paired regions where each region has the characteristic of interest, $x$) is defined as follows:

$$\text{1-1 joins} = 1/2 \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \delta_{ij} x_i x_j.$$

The total number of observed 1-0 joins (i.e., the number of paired regions in which one possesses the characteristic of interest, $x$, and the other one does not) is defined as follows:

$$\text{1-0 joins} = 1/2 \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \delta_{ij}(x_i - x_j)^2.$$

The total number of observed 0-0 joins is found by subtracting the number of 1-1 and 1-0 joins from the total number of joins. The total number of joins in a two-dimensional plane is defined as follows:

$$\text{total joins} = 1/2 \sum_{i=1}^{n} L_i.$$

$L_i$ represents the number of regions joined to the "$i$th" region.

The observed join counts can be tested for significant departures from the randomly expected number of join counts using the Z statistic because join count statistics are asymptotically normally distributed. The first two join count moments can be used as the location (mean) and scale (variance) parameters of the normal distribution when the area in question is a moderately sized lattice. The first two moments of the join counts are listed in Appendix A (Moran, 1948; Cliff and Ord, 1981).

Cliff and Ord (1981) found that the estimated moments provide a reasonable approximation to the normal distribution parameters when the sample sizes are at least equal to 50 (25 joins). These moments can be calculated under two assumptions:

- free sampling (sampling with replacement), in which the individual units are independently assigned the characteristic of interest, with probability $p$; and
- nonfree sampling (sampling without replacement), in which each unit has the same probability, *a priori*, of being assigned the characteristic of interest, but the assignments are subject to the overall constraint that there are $n_1$ units assigned with the characteristic and $n_2$ units without the characteristic, and $n = n_1 + n_2$.

The nonfree sampling assumption is appropriate for most ecological cases in which the regions or individuals are

observed with or without the characteristic of interest. Accordingly, the nonfree sampling assumption is correct to use when analyzing error in remotely sensed data because $p$ is usually unknown and the values of $n_1$ and $n_2$ are found after the entire image has been observed.

There are several important points to note in any spatial autocorrelation analysis. First, spatial autocorrelation analyses may be performed for any lattice shape, any direction, and at any lag or separation distance. A lag considers all possible pairs of data values whose locations are separated by a specific distance in a particular direction (Figure 1). Furthermore, autocorrelation may be expected in a particular direction, such as north-south, or in all directions. Finally, lag distances may be predetermined or dynamic (e.g., in order to identify the lag or lags where autocorrelation is no longer prevalent).

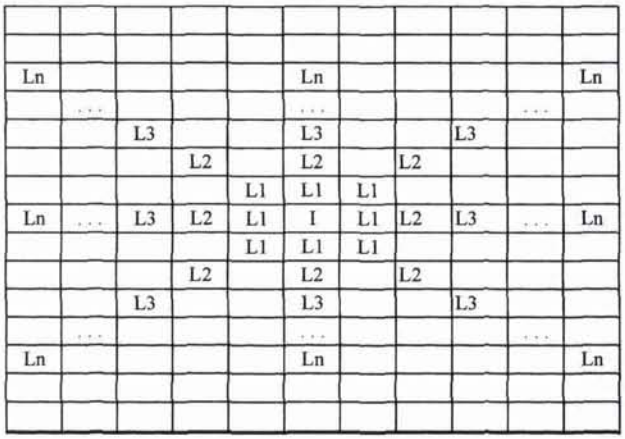


Figure 1. Example representing spatial autocorrelation lags.

#### Landsat Thematic Mapper Data and Reference Data

The main objective of this research was to investigate error in classified satellite imagery using spatial autocorrelation of error analysis. The first step in this process, then, was to obtain classified TM imagery and reference data for comparison. TM imagery was acquired from a subscene of 23 October 1989 imagery (# 5206214513) classified by Schriever and Congalton (1995) with a combination of supervised and unsupervised techniques presented by Chuvieco and Congalton (1988).

The reference forest-cover-type data, on the other hand, were collected on the ground from 1994 through 1995. The data covered some 3600 ha, spanning both public and private land in New Hampshire. The public land area comprised most of Pawtuckaway State Park, a flat to mountainous area with elevations ranging from approximately 76 to 305 m above sea level. The western half of the public land study area includes Mount Pawtuckaway, Middle Mountain, and South Mountain, with steep slopes and many rock outcroppings. The public land is characterized by "wild" forests, protected from development and lightly used by people (Irland, 1982). In contrast, the private land study area, north and adjacent to the public land study area, is mainly rural forests, with a quarter of the area industrial forest and owned primarily by one lumber company. This study area is flat to hilly, with elevations ranging from 76 to 152 m.

The forest inventory of the two areas was initially performed using variable radius plots with a ten-basal-area-factor prism. After the reference data could be confidently estimated visually, transects were walked through stands and visual estimations of the inventory data were recorded in 2-ha minimum mapping units. A compass and pacing in combination with color infrared National High Altitude Photography (NHAP) and topographic maps were used to design and walk the transects. Transects were spaced to minimize overlap within sight distance, yet cover most to all of the stand.

Many variables describing the overstory and understory were inventoried, but overstory species composition was the primary variable used to classify the forest stands. The classification system used for both the TM and reference data was a slightly modified version of the forest-cover-type classification system used by the Society of American Foresters (Eyre, 1980). A detailed description of the classification system is given by Schriever and Congalton (1995). When the forest inventory was complete, it was entered into a database management system and a GIS to create a reference forest-cover-type map of each study area.

#### Forest-Cover-Type Maps Analysis

Analysis of variance (ANOVA) techniques, a chi-square analysis, basic descriptive statistics, and visual inspection were used to compare the study area reference forest-cover-type maps in order to aid in the interpretation of the spatial autocorrelation results. Prior to the ANOVA, reference forest polygon sizes were transformed (ln[polygon size + 1]) based upon the results of a previous exploratory data analysis. In addition, spatial diversity of the forest-cover types in each reference forest-cover-type map was determined with calculations of interspersion. Interspersion was calculated with a modified portion of WILDINT2, a FORTRAN program created by Mead *et al.* (1981). An ANOVA was then applied to test for significant differences in spatial diversity between the forest-cover-type maps.

#### Spatial Autocorrelation of Error Analysis

After comparing the reference forest-cover-type maps, the spatial autocorrelation of error analysis was performed. Prior to the analysis, the classified satellite imagery was "smoothed" with a 3 by 3 low-pass filter in order to more closely match the 2-ha minimum mapping unit of the reference data. Difference images—binary representations of error—were then created to show the patterns of error in the classified TM imagery for both study areas. Next, spatial autocorrelation of error was measured and tested in three directional cases and all valid lag distances using 1-0 join count statistics and the difference images. Finally, the spatial autocorrelation of error was represented with correlograms, and the global significance of these correlograms was tested.

##### *Difference Images*

The classified TM image was registered to the reference forest-cover-type maps (RMS error = 0.30 pixels), which had been registered to the New Hampshire State Plane Coordinate System (North American Datum 1983; RMS error of the private and public lands study areas = 2.78 and 1.94 m, respectively). Each reference forest-cover-type map was compared to the classified TM imagery, pixel by pixel. Difference images were created for each study area to show the agreement and disagreement between the reference forest-cover-type maps and the classified satellite imagery. Each difference image was comprised of a matrix of ones and zeros represented by black and white, respectively, where ones represent error and zeros represent no

error. An ANOVA, basic descriptive statistics, and visual inspection were then used to investigate the structure of error/no error (difference-type) polygons in the difference images. Based upon preliminary exploratory data analysis results, the difference-type polygon sizes were ranked prior to the ANOVA. The investigation into the structure of the difference-type polygons was also anticipated to aid in the interpretation of the spatial autocorrelation results.

*Join Count Statistics*

Following the examination of the difference-type polygons, join count statistics under the nonfree sampling assumption were applied in this spatial autocorrelation of error analysis. The difference images were tested for randomly distributed error using a Z statistic, with the following null hypotheses:

$H_O$ (1) error has an equal probability of occurrence in each pixel of the classified satellite imagery or difference image,

$H_O$ (2) error in one pixel is independent of error in another pixel.

The 1-0 join count statistics were used in this analysis for their high asymptotic efficiency. Cliff and Ord (1981) determined that under nonfree sampling the asymptotic relative efficiency of the 1-0 join count statistics is superior to the 0-0 or 1-1 join count statistics when the numbers of joins per cell (pixel) are not equal. The difference images in this study did not have an equal number of joins per pixel, so using 1-0 join count statistics was further justified. In addition, the asymptotic efficiencies of the 1-0, 1-1, and 0-0 join counts are the highest for nonfree sampling when the proportion of the characteristic of interest is close to or equal to 0.5. If the 1-1 or 0-0 join counts are used and the proportion of the characteristic of interest is not close to 0.5, then the join counts for the proportion of the characteristic greater than 0.5 should be used because they will be more normally distributed and the efficiency of the test statistic will be greater. For example, if the proportion of ones (errors) in a binary population is 0.3 and the proportion of zeros (no error) is 0.7, then the 0-0 join count statistics should be used. The public and private lands difference images had 0.68 and 0.69 proportions of error (ones), respectively. Therefore, the 1-1 join count statistics could have been applied. Nonetheless, the higher asymptotic efficiency of the 1-0 join count statistics was preferred.

The first two moments of the 1-0 join count statistics (Appendix A) were used to calculate measures of spatial autocorrelation and to test whether spatial autocorrelation was significant at the 95 percent confidence level, using the Z statistic. Spatial autocorrelation is significant if the absolute value of the observed Z statistic is greater than the critical Z statistic value at a specified confidence level. Positive spatial autocorrelation occurs if the Z statistic for the 1-0 joins is significant and negative. Negative spatial autocorrelation occurs if the 1-0 joins Z statistic is significant and positive. Positive spatial autocorrelation indicates clumped or blocky error, while negative spatial autocorrelation indicates uniformly distributed error.

Measures of spatial autocorrelation and significance tests were calculated for three directional cases using a FORTRAN program called SAC, a modified version of Congalton's (1984) FORTRAN program, AUTOCOR. The directional cases corresponded to the directions that chess pieces move—specifically, in the directions that the rook, bishop, and queen move. In this analysis, the rook directional case included east-west and north-south directions and the bishop directional case included diagonal directions. In addition, the queen directional case included east-west, north-south, and diagonal directions.

In combination with the different directional cases, the spatial autocorrelation measures and tests of significance were calculated for the statistically valid lag distances which contained at least 50 samples. In this analysis, each "family" of tests consisted of those tests performed for a particular directional case and study area. The number of tests per family equaled the number of lags tested for a specific directional case and study area. The Bonferroni procedure was applied to control the family-wise error rate ($\alpha' = 0.05$). Dunn (1961) showed that the family-wise error rate will be less than $\alpha$ using the Bonferroni procedure in which each test is performed at the $\alpha' = (\alpha/\#$ tests) level. Using a 0.05 family-wise error rate and the Bonferroni procedure, then one would expect to detect one or more false significant tests per family less than 5 percent of the time. The Dunn-Šidák and Holm's sequentially rejective Bonferroni tests were also applied to the lags. These tests not only control family-wise error rates, but also offer more power (Kirk, 1995). Finally, each lag was also tested at the uncorrected or per-lag $\alpha$ level of 0.05 as a comparison to the more conservative Bonferroni procedure.

*Correlograms*

Finally, the spatial autocorrelation of error was represented with correlograms, and the global significance of these correlograms was tested. Correlograms show spatial autocorrelation as a function of lag distance and allow spatial autocorrelation at different lags to be compared and investigated (Odland, 1988). Although uniform lag distances are usually the exception and not the rule in most spatial autocorrelation studies (Odland, 1988), lag distance relations between pairs of first-order, second-order, . . ., "$n$th"-order neighbors in this study were uniform because of the equal-size square pixels of the satellite imagery. Pixel lag distances associated with satellite imagery could also be converted to ground distances, making the correlogram analysis more meaningful.

Correlograms represent the results of multiple autocorrelation measures and tests performed simultaneously. Therefore, a global test of significance is required prior to examining the autocorrelation represented by the correlogram. Oden (1984) presented several alternatives for testing the global significance of correlograms, including the Bonferroni test and the Q test. The least computationally intensive method, the Bonferroni procedure, was chosen to test each correlogram. Correlograms were globally significant if at least one lag showed significant autocorrelation at the $\alpha' = (\alpha/\#$ lags or tests) level. Once global significance was determined, subsequent analyses of the patterns represented in the correlograms were performed. The Z statistic was used as the measure of spatial autocorrelation in the correlograms, and the Bonferroni family-wise and per-lag critical values were displayed to indicate the significance of autocorrelation at particular lags of interest. The lag unit distance equaled one pixel length (28.5 m). Visual inspection of the difference images in conjunction with dimensional measurements of the difference type polygons and distances between prevalent error and no error regions were also used to interpret the correlograms. Furthermore, inferences about the fundamental processes generating the distribution of error could be made by observing the shape of the correlogram.

Various shapes of spatial correlograms may indicate different associated spatial patterns of the characteristic of interest (Sokal and Oden, 1978a; Odland, 1988; Legendre and Fortin, 1989). For example, patches of homogeneous values exhibit gradually decreasing positive autocorrelation as lag distances become longer until no significant autocorrelation occurs. A regular gradient of values exhibits gradually decreasing positive autocorrelation and then negative autocorrelation as lag distances become longer. Sokal and Oden (1978b) and Legendre and Fortin (1989) explained that even though specific patterns lead to characteristic correlograms, however, different patterns may not result in different correlograms. In addition,

correlograms may indicate a particular spatial pattern associated with the characteristic of interest, but no formal statistical tests are available to indicate whether a correlogram is diagnostic of a unique pattern, like regular gradients (Odland, 1988). Finally, correlograms are affected not only by the distribution pattern of the characteristic of interest, but also by the shape of the study area. For example, if a study area is square, then the highest-order lags are only associated with the opposite corners of the study area. As a result, higher-order autocorrelations are affected by directional patterns not present in the lower-order autocorrelations.

## Results

### Forest-Cover-Type Map Comparisons

The mean transformed reference forest polygon size (ln[polygon size + 1]) was greater in the private lands study area (mean = 1.7) than in the public lands study area (mean = 1.5) ($\alpha$ = 0.05, P-value = 0.001, partial omega squared = 0.0089). The overall mean raw forest polygon sizes were 3.98 and 3.46 ha in the private and public lands study areas, respectively. The frequency of occurrences of the reference forest-cover types also differed between the study areas ($\alpha$ = 0.05, P-value = 0.003, Cramer's $V$ = 0.162). However, mean reference forest-cover-type spatial diversity (interspersion) did not differ between the study areas ($\alpha$ = 0.05, P-value = 0.841). These results were supported by visual inspection of the reference forest-cover-type maps, which indicated that the spatial distribution of species within each study area differed considerably.

### Difference Images

The population summary statistics for the private and public lands difference images were very similar. The mean error and variance for the private lands difference image were 0.69 and 0.0030, respectively. The mean error and variance for the public lands difference image were 0.68 and 0.0032, respectively. The results of the ANOVA did not indicate any significant differences between the ranked difference type polygon sizes in the two study areas ($\alpha$ = 0.05, P-value = 0.597). The overall mean raw difference type polygon sizes were 2.77 and 2.57 ha in the private and public lands study areas, respectively.

In addition to the ANOVA results, visual inspection showed that both difference images exhibited periodicities of error with many large contiguous blocks of error (Figures 2a and 2b). However, the error in the private lands difference image was more staggered as opposed to the more concentrated error in the public lands difference image. Furthermore, both difference images showed contiguous blocks of no error, with more prevalent linear trends of no error occurring in the public lands difference image.

### Correlograms and Difference Images

All the correlograms for the public and private lands difference images were globally significant, displaying significant patterns of spatial autocorrelation of error at the $\alpha'$ = (0.05/# lags) level. Figures 3a and 3b show the queen directional case correlogram for each study area. The family-wise ($\alpha'$) and per-lag ($\alpha$) critical levels are displayed; however, the following correlogram descriptions are presented with respect to the family-wise critical level.

Randomly distributed error first occurred at shorter lag distances in the private lands difference image for all three directional cases (lag 15 and 23 for the private and public lands queen case correlograms, respectively). In addition, the correlograms of the private lands difference image included less valid lag distances. All three correlograms for the private lands difference image indicated that clumped spatial autocorrelation of error occurred initially, followed by alternating uniform and clumped spatial autocorrelation with intervals of randomly distributed error. The staggering of error versus no error throughout the difference image was associated with these generally alternating intervals of opposing spatial autocorrelation.

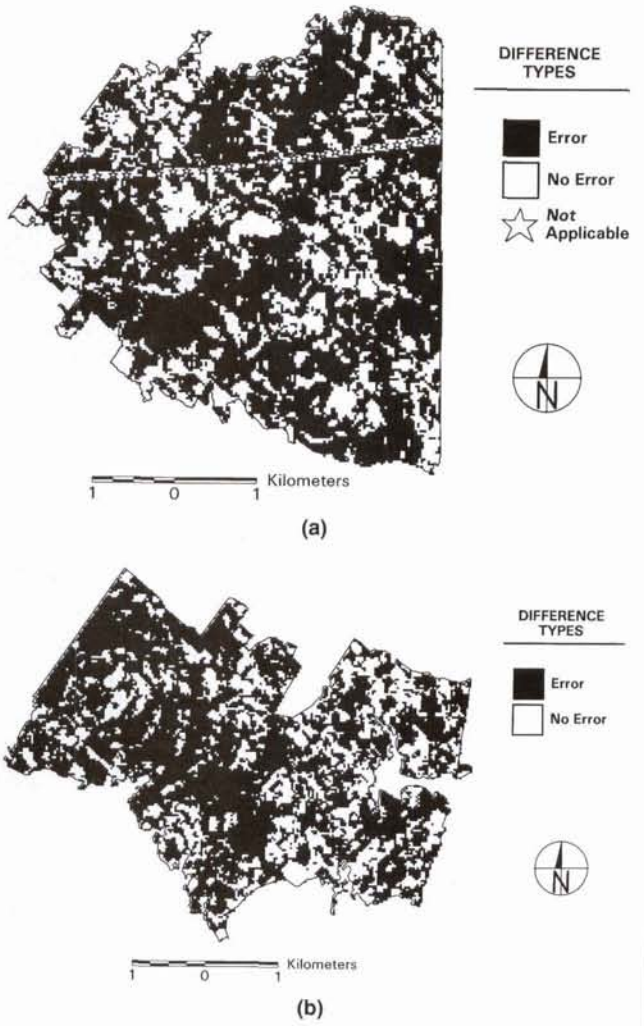


Figure 2. (a) Private lands forest-cover-type difference image. (b) Public lands forest-cover-type difference image.

In contrast, all of the public lands correlograms showed that a clumped pattern of error occurred initially followed by uniformly or randomly distributed error. In the queen directional case, the first interval of uniformly distributed error coincided with the comparisons between the large block of error in the south-central region and the no-error regions throughout the rest of the difference image at the lag distances within the interval. The second interval of uniformly distributed error corresponded with the comparisons between the error in the western one-third of the difference image and the relatively more accurate northeast and southeast corners.

## Discussion

### Reference Data

The detailed reference forest inventory data stands out as one of the primary highlights in this research. Acquisition of reference data is very time consuming and results in an immense expenditure. Consequently, detailed reference data are uncommon for large contiguous areas. The reference data utilized here enhanced the reliability of the conclusions.

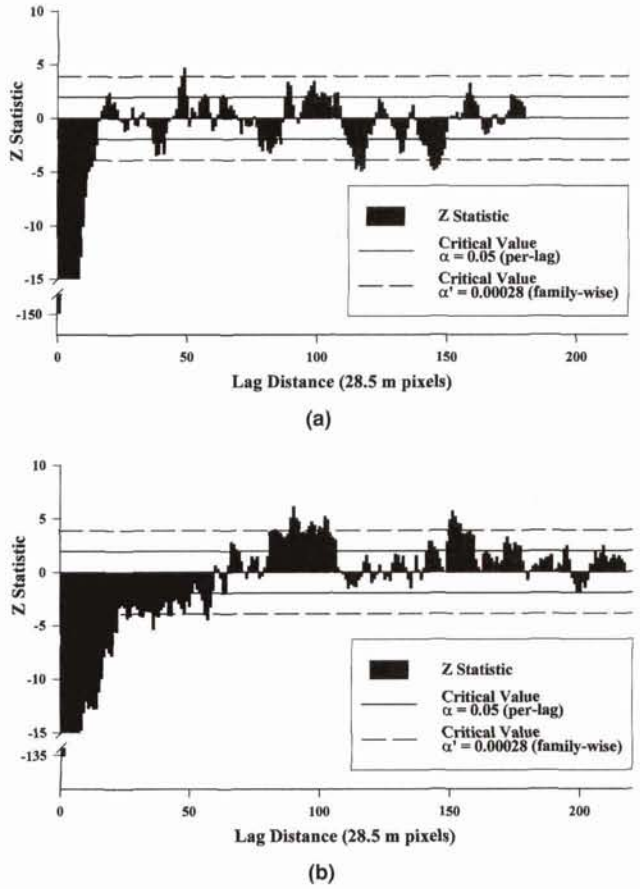


Figure 3. (a) Queen directional case correlogram for the private lands difference image using 1-0 join count statistics. (b) Queen directional case correlogram for the public lands difference image using 1-0 join count statistics.

Walking transects and visually estimating inventory data provided a complete overview of forest stands in less time than required for an adequate sampling using actual hands-on measurements. In addition, the visual ground estimation provided more dependable results than did photointerpretation. Congalton and Green (1993) identified differences between ground and photointerpretations, determining that, even with large-scale photography, the understory cannot be consistently identified in many parts of New England due to the dense overstory canopy closure. This is key, because the understory composition can be a deciding factor when identifying forest-cover types. Indeed, in many instances in this research, the reference forest-cover type was determined while considering the understory composition.

#### Spatial Autocorrelation of Error in the Public versus the Private Lands Study Areas

Each study area exhibited different patterns of error. Concentrated blocks of error were more pronounced in the public lands study area relative to the more evenly distributed error observed in the private lands study area. The concentrated blocks of error caused the initial positive and then negative spatial autocorrelation. In contrast, the evenly distributed error caused the generally alternating positive and negative spatial autocorrelation. Sokal and Oden (1978a) investigated correlograms generated from artificial patterns of joins, concluding that regular alternations between negative and positive autocorrelations are expected for some surface patterns where the values alternate from high to low. This was the case in the private lands difference image. Furthermore, the concentrated clumps of error in the public lands study area caused the first randomly distributed error to occur at relatively longer lag distances.

Differences in the forest-cover-type patterns of the study areas contributed to the differences in spatial autocorrelation of error between the study areas. Differences in the physiography between the study areas may further have added to the different patterns of error. Much of the error in the public lands study area coincided with mountainous terrain, where the TM imagery indicated dark shadows on the western aspects and very bright reflections on the eastern aspects. Congalton (1988a) suspected that topography, specifically shadow caused by mountains, contributed to the spatial autocorrelation of error observed in a forested environment difference image. Similarly, Labovitz *et al.* (1982) determined, in their investigation of spatial autocorrelation of spectral reflectance values in TMS data, that physiography was more important than land-cover class in explaining autocorrelation effects. No other factors could be identified as possible causes contributing to the different patterns of error in the two study areas.

The ANOVA, chi-square analysis, and basic descriptive statistics identified some differences between the study areas. Still, visual inspection of the reference forest-cover-type maps and difference images was the most efficient and effective method used to explain the differences in spatial autocorrelation of error. For example, the private lands reference forest-cover-type map had slightly larger polygon sizes on average; however, the polygons of the public lands reference forest-cover-type map were more regularly shaped with fewer small island polygons. This contributed to the greater concentrations of error in the public lands difference image. In addition, the insignificant differences in spatial diversity and mean difference type polygon size were poor indicators of the differences in spatial autocorrelation of error. The chi-square analysis did identify contrasting relative frequencies of forest-cover types between the study areas, but the critical factor was the difference in the forest-cover-type spatial distributions between the study areas.

#### Multiple Spatial Autocorrelation Tests

The Bonferroni procedure was used to control the family-wise error rate. The alternative Dunn-Šidák and Holm's sequentially rejective Bonferroni test procedures were also applied to the data. However, increases in significant lags were zero to few in each family of tests, revealing no strikingly apparent differences in the correlograms. Regardless, the correlograms were globally significant. That is, there were significant factors causing different patterns of error to occur in each study area. Factors such as those previously mentioned (e.g., physiography) may cause more or less significantly autocorrelated lags in other cases.

Previous researchers have indicated significant spatial autocorrelation at the uncorrected or per-lag alpha level after determining global significance (e.g., Sokal and Oden, 1978a; Legendre and Fortin, 1989). The per-lag critical level was shown in this study for comparison purposes. The preference in this analysis was to control the family-wise error rate at 0.05. While some may regard this as a conservative approach possibly resulting in more Type II errors, fewer Type I errors may also occur.

Finally, global significance may have been overestimated. While six correlograms were tested for global significance, and each was declared significant, the probability of falsely declaring one or more globally significant correlograms was equal to 0.26. This was determined using the following formula from Kirk (1995) where $C$ is the number of independent tests:

probability of one or more Type I errors = $1 - (1 - \alpha)^C$.

#### Sampling and Spatial Autocorrelation of Error

Systematic sampling has been shown to perform poorly when periodicity of error is extreme. However, it performs well when the correlogram exhibits a monotonically decreasing function. Berry and Baker (1968) noted that systematic sampling had the greatest relative efficiency (unbiased with the lowest relative variance) when spatial autocorrelation declined monotonically with increasing lag distance. Matérn (1960) also discovered that systematic sampling with one unit per stratum was superior to simple or stratified random sampling when the correlogram declined monotonically. The correlograms in this study, however, did not exhibit monotonically decreasing functions. Instead, extreme periodicities of error were expressed. Therefore, systematic sampling should be used with caution. Indeed, systematic sampling may not perform adequately when estimating the accuracy of the classified TM imagery in this study. Systematic sampling should work adequately if samples are spaced at lag distances exhibiting no spatial autocorrelation. However, analysts usually do not have correlograms to indicate the appropriate spacing. As shown, the choice of sampling scheme depends partly upon the distribution of the phenomenon being sampled. Nonetheless, if the phenomenon is randomly distributed, then the previously mentioned sampling schemes should all provide unbiased estimates with nearly equal variances (Berry and Baker, 1968).

Given the spatial autocorrelation of error observed in this study, adequate sampling schemes should be identified. The patterns of spatial autocorrelation observed in this study have not been identified in previous remote sensing analyses. In addition, Congalton's (1988a; 1988b) research was the only remote sensing analysis investigating spatial autocorrelation of error and sampling schemes simultaneously. As in Congalton's (1988b) study, a Monte Carlo simulation would be an appropriate method to apply to explore appropriate sampling schemes. If extreme periodicities of error are expected, then an analyst may be able to choose the proper sampling scheme with guidance from this future research.

#### Spatial Autocorrelation of Error in TM versus MSS Imagery

Would the spatial autocorrelation of error have differed in this study if an MSS image was used instead of a TM image? Significantly higher accuracies are commonly achieved when the increased spectral and radiometric properties of the TM imagery are applied (e.g., Williams *et al.*, 1984), yet the increased spatial resolution of the TM imagery causes accuracies to fluctuate depending upon the within-class variability and percentage of boundary pixels within a scene (e.g., Irons *et al.*, 1984). The lower spectral and radiometric resolutions of the MSS imagery may well have decreased accuracy in this study. On the other hand, the lower spatial resolution of the MSS imagery could presumably have increased accuracy because the study areas in the TM imagery contained a high degree of within-class variability. Nonetheless, the TM imagery would likely have yielded the greatest net benefit with the highest accuracies.

The cumulative effect of factors such as spectral, radiometric, and spatial resolution on spatial autocorrelation of error is difficult to predict. For example, the lower spatial resolution of the MSS imagery decreases the number of samples in a given area, possibly lowering the degree of spatial autocorrelation, clumped or uniform, for a particular lag distance. The difference in spatial autocorrelation of error would finally depend on the relative improvement of the TM classification over the MSS classification. Future work should include the classification of TM and MSS imagery covering the same area while holding other variables constant. An alternative would involve degrading TM imagery to the various spectral, radiometric, and spatial resolutions of the MSS imagery and comparing the effects of these factors on spatial autocorrelation of error.

## Conclusions

No broad conclusions can be made about the differences in spatial autocorrelation of error for TM data in private versus public lands. Yet differences in spatial autocorrelation of error clearly can occur in TM data. Therefore, systematic sampling may not always be appropriate to use with spatially autocorrelated data.

Many factors can affect accuracy and spatial autocorrelation of error. Preprocessing and classification techniques, temporal and cover type variability, climatic phenomena, physiography, and resolution are some examples. Therefore, generalizations describing spatial autocorrelation for a particular type of environment or imagery should be interpreted with caution. Each case should be judged separately while considering all factors. Furthermore, visual inspection of ancillary data such as maps and remotely sensed images can be especially supportive when determining spatial autocorrelation of error and appropriate sampling schemes. More research investigating factors contributing to spatially autocorrelated error will be the key to knowing the best type of prevention and/or compensation.

## Acknowledgments

Special thanks go to Alan Fried, Robb Macleod, Dan Fehringer, Jennifer Ashby, Dr. William Stine, and Dr. John Aber for assistance throughout this study. Dr. Stephen Stehman also provided valuable advice and comments. In addition, the authors acknowledge the funding provided by McIntire-Stennis Grant #MS-33 through the University of New Hampshire.

## References

Berry, B.J.L., and A.M. Baker, 1968. Geographic sampling, *Spatial Analysis: A Reader in Statistical Geography*, (B.J.L. Berry and D.F. Marble, editors), Prentice Hall, Englewood Cliffs, N.J., pp. 91–100.

Campbell, J.B., 1981. Spatial correlation effects upon accuracy of supervised classification of land cover, *Photogrammetric Engineering & Remote Sensing*, 47(3):355–363.

Chuvieco, E., and R.G. Congalton, 1988. Using cluster analysis to improve the selection of training statistics in classifying remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 56(9):1275–1281.

Cliff, A.D., and J.K. Ord., 1981. *Spatial Processes: Models and Applications*, Pion Limited, London, England, 266 p.

Congalton, R., 1984. *A Comparison of Five Sampling Schemes Used in Assessing the Accuracy of Land Cover/Land Use Maps Derived from Remotely Sensed Data*, Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 147 p.

———, 1988a. Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 54(5):587–592.

———, 1988b. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 54(5):593–600.

Congalton, R.G., and K. Green, 1993. A practical look at the sources of confusion in error matrix generation, *Photogrammetric Engineering & Remote Sensing*, 59(5):641–644.

Craig, R.G., and M.L. Labovitz, 1980. Sources of variation in Landsat autocorrelation, *Proceedings of the Fourteenth International Symposium on Remote Sensing of Environment*, San Jose, Costa Rica, pp. 1755–1767.

Dunn, O.J., 1961. Multiple comparisons among means, *Journal of the American Statistical Association*, 56:52–64.

Eyre, F.H. (editor), 1980. *Forest Cover Types of the United States and Canada*, Society of American Foresters, Washington, D.C., 148 p.

Irland, L.C., 1982. *Wildlands and Woodlots - The Story of New England's Forests*, University Press of New England, Hanover, New Hampshire, 217 p.

Irons, J.R., B.L. Markham, R.F. Nelson, D.L. Toll, D.L. Williams, R.S. Latty, and M.L. Stauffer, 1984. The effects of spatial resolution on the classification of Thematic Mapper data, *Int. J. of Remote Sensing*, 6(8):1385–1403.

Kirk, R.E., 1995. *Experimental Design: Procedures for the Behavioral Sciences*, Brooks/Cole, Pacific Grove, California, 921 p.

Labovitz, M.L., D.L. Toll, and R.E. Kennard, 1982. Preliminary evidence for the influence of physiography and scale upon the autocorrelation function of remotely sensed data, *Int. J. of Remote Sensing*, 3(1):13–30.

Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74(6):1659–1673.

Legendre, P., and M.-J. Fortin, 1989. Spatial pattern and ecological analysis, *Vegetatio*, 80:107–138.

Matérn, B., 1960. Spatial variation. *Meddelanden Fran Statens Skogsforskningsinstitut*, 49(5):1–144.

Mead, R.A., T.L. Sharik, S.P. Prisley, and J.T. Heinen, 1981. *A Computerized Spatial Analysis System for Assessing Wildlife Habitat from Vegetative Maps*, USDA Forest Service, Nationwide Forestry Applications Program Renewable Resources Inventory Project, Blacksburg, Virginia, pp. 66–85.

Moran, P.A.P., 1948. The interpretation of statistical maps, *J. Royal Stat. Soc., Series B*, 37:243–251.

Oden, N.L., 1984. Assessing the significance of a spatial correlogram, *Geographical Analysis*, 6(1):1–16.

Odland, J., 1988. *Spatial Autocorrelation*, Sage, Newbury Park, California, 178 p.

Schriever, J.R., and R.G. Congalton, 1995. Evaluating seasonal variability as an aid to cover-type mapping from Landsat Thematic Mapper data in the Northeast, *Photogrammetric Engineering & Remote Sensing*, 61(3):321–327.

Sokal, R.R., and N.L. Oden, 1978a. Spatial autocorrelation in biology 1: Methodology, *Biological J. of the Linnean Soc.*, 10:199–228.

———, 1978b. Spatial autocorrelation in biology 2: Some biological applications of evolutionary and ecological interest, *Biological J. of the Linnean Soc.*, 10:229–249.

Tubbs, J.D., and W.A. Coberly, 1978. Spatial correlation and its effect upon classification results in Landsat, *Proceedings of the Twelfth International Symposium on Remote Sensing of Environment*, Environmental Research Institute of Michigan, Ann Arbor, Michigan, pp. 775–781.

Williams, D.L, J.R. Irons, B.L. Markham, R.F. Nelson, D.L. Toll, R.S. Latty, and M.L. Stauffer, 1984. A statistical evaluation of the advantages of Landsat Thematic Mapper data in comparison to Multispectral Scanner data, *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(3):294–301.

## Appendix A. Join Count Statistics

Moments of the Join Count Statistics (Moran, 1948):

Moments of the 1-1 join count statistic:

$$u'_1\,(1\text{-}1) = \frac{An_1^{(2)}}{n^{(2)}}$$

$$u_2\,(1\text{-}1) = \frac{An_1^{(2)}}{n^{(2)}} + \frac{2Dn_1^{(2)}}{n^{(3)}} + \frac{[A(A-1)-2D]n_1^{(4)}}{n^{(4)}} - \left[\frac{An_1^{(2)}}{n^{(2)}}\right]^2$$

Moments of the 0-0 join count statistic:

$$u'_1\,(0\text{-}0) = \frac{An_2^{(2)}}{n^{(2)}}$$

$$u_2\,(0\text{-}0) = \frac{An_2^{(2)}}{n^{(2)}} + \frac{2Dn_2^{(3)}}{n^{(3)}} + \frac{[A(A-1)-2D]n_2^{(4)}}{n^{(4)}} - \left[\frac{An_2^{(2)}}{n^{(2)}}\right]^2$$

Moments of the 1-0 join count statistic:

$$u'_1\,(1\text{-}0) = \frac{2An_1n_2}{n^{(2)}}$$

$$u_2\,(1\text{-}0) = \frac{2An_1n_2}{n^{(2)}} + \frac{4[A(A-1)-2D]n_1^{(2)}n_2^{(2)}}{n^{(4)}} + \frac{2Dn_1n_2\,(n_1+n_2-2)}{n^{(3)}} - 4\left[\frac{An_1n_2}{n^{(2)}}\right]^2$$

Notation of the Moments (Moran, 1948):

$u_1'(x)$ = the first moment of $x$ about the origin

$u_2(x)$ = the second moment of $x$ about the mean

$n^{(i)} = n(n-1)\ldots(n-i-1)$

$n$ = the total number of regions in the population

$n_1$ = the total number of regions in the population with the characteristic of interest

$n_2$ = the total number of regions in the population without the characteristic of interest

$$A = 1/2\sum_{i=1}^{n} L_i$$

$$D = 1/2\sum_{i=1}^{n} L_i(L_i-1)$$

$L_i$ = the number of regions joined to the "$i$th" region

(Received 30 November 1999; accepted 23 March 2000; revised 16 May 2000)